

MULTILEVEL ENSEMBLE KALMAN FILTERING FOR SPATIALLY EXTENDED MODELS

ALEXEY CHERNOV^{††}, HÅKON HOEL^{*}, KODY J. H. LAW[†], FABIO NOBILE^{°°},
AND RAUL TEMPONE[°]

ABSTRACT. This work embeds a multilevel Monte Carlo (MLMC) sampling strategy into the Monte Carlo step of the ensemble Kalman filter (EnKF), thereby yielding a multilevel ensemble Kalman filter (MLEnKF) which has provably superior asymptotic cost to a given accuracy level. The development of MLEnKF for finite-dimensional state-spaces in the work [20] is here extended to models with infinite-dimensional state-spaces in the form of spatial fields. A concrete example is given to illustrate the results.

Key words: Monte Carlo, multilevel, filtering, Kalman filter, ensemble Kalman filter. **AMS subject classification:** 65C30, 65Y20,

1. INTRODUCTION

Filtering refers to the sequential estimation of the state v and/or parameters p of a system through sequential incorporation of online data y . The most complete estimation of the state v_n at time n is given by its probability distribution conditional on the observations up to the given time $\mathbb{P}(dv_n|y_1, \dots, y_n)$ [23, 1]. For linear Gaussian systems the analytical solution may be given in closed form, via update formulae for the mean and covariance known as the Kalman filter [24]. However, in general there is no closed form solution. One must therefore resort to either algorithms which approximate the probabilistic solution by leveraging ideas from control theory [25, 23], or Monte Carlo methods to approximate the filtering distribution itself [1, 11, 7]. The ensemble Kalman filter (EnKF) [5, 13] combines elements of both approaches. In the linear Gaussian case it converges to the Kalman filter solution [37], and even in the nonlinear case, under suitable assumptions it converges [33, 32] to a limit which, for a single update,

^{††}Institute for Mathematics, Carl von Ossietzky University Oldenburg, Germany (alexey.chernov@uni-oldenburg.de).

^{*}Department of Mathematics, University of Oslo, Norway (haakonah@math.uio.no).

[†]Computer Science and Mathematics Division, Oak Ridge National Laboratory (lawkj@ornl.gov).

^{°°}Mathematics Institute of Computational Science and Engineering, École polytechnique fédérale de Lausanne, Switzerland (fabio.nobile@epfl.ch).

[°]Applied Mathematics and Computational Sciences, KAUST, Thuwal, Saudi Arabia (raul.tempone@kaust.edu.sa).

is optimal among those which incorporate the data linearly [32, 34, 39]. In the case of spatially extended models approximated on a numerical grid, the state space itself may become very high-dimensional and even the linear solves may become intractable. Therefore, one may be inclined to use the EnKF filter even for linear Gaussian problems in which the solution is intractable despite being given in closed form on paper by the Kalman filter.

Herein the underlying problem will admit a hierarchy of approximations with cost inversely proportional to accuracy, and it will be necessary to approximate the target for a single prediction step. It has been proposed to use a multilevel identity to optimize the work required to achieve a certain total error level in the Monte Carlo approximation of such random fields [18]. See [15] for a recent review of multilevel Monte Carlo (MLMC). Very recently, a number of works have emerged which extend the MLMC framework to the context of Monte Carlo algorithms designed for Bayesian inference. Examples include Markov chain Monte Carlo [28, 19], sequential Monte Carlo samplers [2, 22, 8], particle filters [21, 16], and EnKF [20]. The filtering papers [21, 16, 20] thusfar all consider only finite-dimensional SDE forward models, with the approximation error as arising from time discretization.

The present work considers the extension of the multilevel EnKF (MLEnKF) [20] to spatially extended models. The infinite-dimensional case was considered in the context of the square root EnKF in [29]. As in that work, we will require that the limiting covariance is trace-class. It was mentioned above that the limiting EnKF distribution, the so-called mean-field EnKF (MFEnKF), is in general not the Bayesian posterior filtering distribution and has a fixed bias. The error of the EnKF approximation may be decomposed into MC error and this Gaussian bias as shown in [32]. According to folklore, small sample sizes are suitable, and it may well be due to minimum error being limited by the bias. Nonetheless, the latter is difficult to quantify and deal with, while the MC error can be controlled and minimized. Unfortunately, scientists are often limited to small ensemble sizes anyway, due to an extremely high-dimensional underlying state space, which is approximating a spatial field. Within the MLEnKF framework developed here, a much smaller MC error can be obtained for the same fixed cost, which will lower the cost requirement for practitioners to ensure that the MC error is commensurate with the bias. Furthermore, it has been shown in [27, 42, 41, 26] that signal tracking stability of EnKF is based on a feedback control mechanism, which can be established for a single member ensemble in 3DVAR [4, 3, 31, 40, 38, 17, 14]. The greater accuracy of EnKF in comparison to 3DVAR [30] is afforded presumably by its use of the ensemble statistics, and the relation to the optimal linear update. Therefore, it is of interest to improve the MC approximation.

The rest of the paper will be organized as follows. In section 2 the notation and problem will be introduced, and the spatial multilevel EnKF (MLEnKF) will be introduced for the first time in sub-section 2.4. In section 3 it is proven that indeed the spatial MLEnKF inherits almost the same favorable

asymptotic “cost-to- ε ” as the standard MLMC for a finite time horizon, and its mean-field limiting distribution is the filtering distribution in the linear and Gaussian case. In section 4 a concrete example will be given to illustrate the theory. Finally, conclusions and future directions are presented in section 5.

2. KALMAN FILTERING

2.1. General set-up. Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a complete probability space, where Ω is the set of events, \mathcal{E} is the sigma algebra of subsets of Ω and \mathbb{P} is the associated probability measure. Let \mathcal{H} be a separable Hilbert space and $L^p(\Omega; \mathcal{H}) = \{u : \Omega \rightarrow \mathcal{H}; \mathbb{E}\|u\|_{\mathcal{H}}^p < \infty\}$, and denote the associated norm $\|u\|_{L^p(\Omega; \mathcal{H})} = (\mathbb{E}\|u\|_{\mathcal{H}}^p)^{1/p}$, or just $\|u\|_p$ where the meaning is clear. Consider the general stochastic signal evolution for the random variables $u_n \in L^p(\Omega; \mathcal{H})$, where,

$$u_{n+1} = \Psi(u_n), \quad (2.1)$$

for $n = 0, 1, \dots, N-1$. In particular, we will be concerned herein with the case in which $\Psi : L^p(\Omega; \mathcal{H}) \rightarrow L^p(\Omega; \mathcal{H})$ is the finite-time evolution of an SPDE or, equivalently, a discrete random mapping (possibly nonlinear) of a spatially extended state given as a random L^p integrable element of the separable Hilbert space \mathcal{H} . Let $\{\phi_k\}_{k=1}^\infty$ be a countable orthonormal basis spanning the Hilbert space \mathcal{H} , so that elements $u \in \mathcal{H}$ admit the representation $u = \sum_{k=1}^\infty u^k \phi_k$, where $u^k = \langle u, \phi_k \rangle_{\mathcal{H}}$. The notation $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and $\cdot \otimes \cdot$ is used to denote the inner and outer products over \mathcal{H} , with the induced norm $\|\cdot\|_{\mathcal{H}} := \langle \cdot, \cdot \rangle_{\mathcal{H}}^{1/2}$, while for finite-dimensional spaces we assign the notation $\mathcal{R}_d = (\mathbb{R}^d, \langle \cdot, \cdot \rangle)$ to denote the Hilbert space with the Euclidean inner product and the induced norm $\|\cdot\|_{\mathcal{R}_d} := \langle \cdot, \cdot \rangle^{1/2}$. Where required, the spatial variable will be denoted with $x, z \in \mathbb{R}^d$ for some $0 < d < \infty$.

Given the history of signal observations

$$y_n = Hu_n + \eta_n, \quad (2.2)$$

with $H : \mathcal{H} \rightarrow \mathbb{R}^m$ linear and η_n are i.i.d. with $\eta_1 \sim N(0, \Gamma)$, $\Gamma \in \mathbb{R}^{m \times m}$ symmetric positive definite, the objective is to track the signal u_n given the observations Y_n where $Y_n = (y_1, y_2, \dots, y_n)$. Notice that under the given assumptions we have a hidden Markov model. That is, the distribution of the random variable we seek to approximate admits the following sequential

structure

$$\begin{aligned}
\mathbb{P}(du_n|Y_n) &= \frac{1}{Z(Y_n)} \mathcal{L}(u_n; y_n) \mathbb{P}(du_n|Y_{n-1}), \\
\mathbb{P}(du_n|Y_{n-1}) &= \int_{u_{n-1} \in \mathcal{H}} \mathbb{P}(du_n|u_{n-1}) \mathbb{P}(du_{n-1}|Y_{n-1}), \\
\mathcal{L}(u_n; y_n) &= \exp\left\{-\frac{1}{2} \|\Gamma^{-1/2}(y_n - Hu_n)\|_{\mathcal{R}_d}^2\right\}, \\
Z(Y_n) &= \int_{u_n \in \mathcal{H}} \mathcal{L}(u_n; y_n) \mathbb{P}(du_n|Y_{n-1}).
\end{aligned} \tag{2.3}$$

It will be assumed that $\Psi(\cdot)$ cannot be evaluated exactly, but rather only approximately, and that there exists a hierarchy of accuracies at which it can be evaluated, each with its associated cost. The explicit dependence on ω will be suppressed where confusion is not possible. For notational simplicity, we will consider the particular case in which the map $\Psi(\cdot)$ does not depend on n . Note that the results easily extend to the non-autonomous case, provided the given assumptions on Ψ are uniform with respect to $\{\Psi_n\}_{n=1}^N$. The specialization is merely for notational convenience. In particular, we will need to denote by $\{\Psi^\ell\}_{\ell=0}^\infty$ a hierarchy of approximations to the solution $\Psi := \Psi^\infty$. First some assumptions must be made.

Assumption 1. *For every $p \geq 2$, the solution operators $\{\Psi^\ell\}_{\ell=0}^\infty$ satisfy the following conditions, for some $0 < c_\Psi < \infty$ depending on Ψ :*

- (i) $\|\Psi^\ell(u) - \Psi^\ell(v)\|_{L^p(\Omega; \mathcal{H})} \leq c_\Psi \|u - v\|_{L^p(\Omega; \mathcal{H})}$,
- (ii) $\|\Psi^\ell(u)\|_{L^p(\Omega; \mathcal{H})}^p \leq c_\Psi (1 + \|u\|_{L^p(\Omega; \mathcal{H})}^p)$.

The covariance matrix of random variables $Z, X \in \mathcal{H}$ will be denoted

$$\text{Cov}[Z, X] := \mathbb{E}[(Z - \mathbb{E}[Z]) \otimes (X - \mathbb{E}[X])],$$

with the shorthand $\text{Cov}[Z] := \text{Cov}[Z, Z]$.

2.2. Some details on Hilbert spaces, Hilbert-Schmidt operators, and Cameron-Martin spaces. Let \mathcal{K}_1 and \mathcal{K}_2 be two separable Hilbert spaces with inner products $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{K}}$ and the induced norms

$$\|u\|_{\mathcal{K}_1} := \langle u, u \rangle_{\mathcal{K}_1}^{1/2}, \quad \text{and} \quad \|u\|_{\mathcal{K}_2} := \langle u, u \rangle_{\mathcal{K}_2}^{1/2}. \tag{2.4}$$

The tensor product of \mathcal{K}_1 and \mathcal{K}_2 is a Hilbert space with the inner product defined by

$$\langle u \otimes v, u' \otimes v' \rangle_{\mathcal{K}_1 \otimes \mathcal{K}_2} = \langle u, u' \rangle_{\mathcal{K}_1} \langle v, v' \rangle_{\mathcal{K}_2} \quad \forall u, u' \in \mathcal{K}_1, \quad \forall v, v' \in \mathcal{K}_2 \tag{2.5}$$

and extended by linearity to finite sums. The tensor product $\mathcal{K}_1 \otimes \mathcal{K}_2$ is the completion of this set with respect to the induced norm $\|\cdot\|_{\mathcal{K}_1 \otimes \mathcal{K}_2}$. It holds that

$$\|u \otimes v\|_{\mathcal{K}_1 \otimes \mathcal{K}_2} = \|u\|_{\mathcal{K}_1} \|v\|_{\mathcal{K}_2}. \tag{2.6}$$

Notice furthermore that every $u \otimes v \in \mathcal{K}_1 \otimes \mathcal{K}_2$ can be identified with a bounded linear mapping

$$T_{u,v} : \mathcal{K}_2^* \rightarrow \mathcal{K}_1 \quad \text{with} \quad T_{u,v}(f) := f(v)u, \text{ for } f \in \mathcal{K}_2^*. \quad (2.7)$$

For two bounded linear operators $A, B : \mathcal{K}_2^* \rightarrow \mathcal{K}_1$ we recall the definition of the Hilbert-Schmidt inner product and the norm

$$\langle A, B \rangle_{HS} = \sum_{k=1}^{\infty} \langle Ae_k^*, Be_k^* \rangle_{\mathcal{K}_1}, \quad |A|_{HS} = \langle A, A \rangle_{HS}^{1/2}, \quad (2.8)$$

where $\{e_k^*\}_{k=1}^{\infty}$ is any complete orthonormal sequence in \mathcal{K}_2^* . A bounded linear operator $A : \mathcal{K}_2^* \rightarrow \mathcal{K}_1$ is called a Hilbert-Schmidt operator if $|A|_{HS} < \infty$ and $HS(\mathcal{K}_2^*, \mathcal{K}_1)$ is the space of all such operators. In view of (2.7) we observe

$$\begin{aligned} |T_{u,v}|_{HS}^2 &= \sum_{k=1}^{\infty} \langle e_k^*(v)u, e_k^*(v)u \rangle_{\mathcal{K}_1} \\ &= \|u\|_{\mathcal{K}_2}^2 \sum_{k=1}^{\infty} |e_k^*(v)|^2 = \|u\|_{\mathcal{K}_1}^2 \|v\|_{\mathcal{K}_2}^2 = \|u \otimes v\|_{\mathcal{K}_1 \otimes \mathcal{K}_2}, \end{aligned}$$

and therefore the tensor product space $\mathcal{K}_1 \otimes \mathcal{K}_2$ is isometrically isomorphic to $HS(\mathcal{K}_2^*, \mathcal{K}_1)$ (and to $HS(\mathcal{K}_2, \mathcal{K}_1)$ by the Riesz representation theorem). For an element $A \in \mathcal{K}_1 \otimes \mathcal{K}_2$ we identify the norms

$$\|A\|_{\mathcal{K}_1 \otimes \mathcal{K}_2} = |A|_{HS}. \quad (2.9)$$

Consider the Gaussian random variable $u \sim \mu_0 := N(0, C)$. Provided the spectrum of C is trace-class, then it has an eigen-basis which is orthonormal with respect to \mathcal{H} , in the sense that $C\phi_k = \lambda_k \phi_k$, $\langle \phi_j, \phi_k \rangle_{\mathcal{H}} = \delta_{j,k}$, and $\sum_{k=0}^{\infty} \lambda_k < \infty$. It is easy to see that $u \in \mathcal{H}$ μ_0 -almost surely. The space $E := \{v \in \mathcal{H}; \|C^{-1/2}v\|_{\mathcal{H}} < \infty\}$ is known as the Cameron-Martin space, and it is also clear, by Kolmogorov's three series theorem, cf. [6], that $u \sim \mu_0 \Rightarrow u \notin E$ almost surely. In fact, $E \subset \mathcal{H} \subset E^*$, where E^* denotes the dual of E wrt the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, and $C : E^* \rightarrow E$.

Proposition 1. *If $u \in L^2(\Omega; \mathcal{H})$ then $C := \mathbb{E}[(u - \mathbb{E}[u]) \otimes (u - \mathbb{E}[u])] \in \mathcal{H} \otimes \mathcal{H}$. Furthermore, $C : \mathcal{H} \rightarrow E^2$, where $E^2 := \{v \in \mathcal{H}; \|C^{-1}v\|_{\mathcal{H}} < \infty\} \subset E$.*

Proof. Notice that $\|\mathbb{E}[u]\|_{\mathcal{H}}^2 \leq \mathbb{E}[\|u\|_{\mathcal{H}}^2]$ by Jensen's inequality, so $\mathbb{E}[u] \in \mathcal{H}$, since $u \in L^2(\Omega; \mathcal{H})$. Without loss of generality let $\mathbb{E}[u] = 0$. Noting that $\text{Tr}(\mathbb{E}[u \otimes u]) = \mathbb{E}[\|u\|_{\mathcal{H}}^2]$ provides the first claim. The second part is obvious since $v = C^{-1}(Cv)$. \square

2.3. EnKF. EnKF uses an ensemble of particles to estimate means and covariance matrices appearing in the Kalman filter, however the framework can be generalized to non-Gaussian models. Let $v_{n,i}$, $\hat{v}_{n,i}$ respectively denote the prediction and update of the i -th particle at simulation time n .

One EnKF two-step transition consists of the propagation of an ensemble $\{\hat{v}_{n,i}\}_{i=1}^M \mapsto \{\hat{v}_{n+1,i}\}_{i=1}^M$.¹ This procedure consists nonetheless in the predict and update steps. In the predict step, M particle paths are computed over one interval, i.e.,

$$v_{n+1}(\omega_i) = \Psi(\hat{v}_n(\omega_i), \omega_i) \quad (2.10)$$

for $i = 1, \dots, M$, where $v_n(\omega_i) := v_{n,i}$ denotes a realization corresponding to the event sample ω_i of the random variable $v_n : \Omega \rightarrow \mathcal{H}$, and $\Psi(\cdot, \omega_i)$ signifies the corresponding realization of the map for a given initial condition. Indeed the notation for random variable realizations, e.g. $\xi_{n,i}$ and $\xi_n(\omega_i)$, will be used interchangeably where confusion is not possible. The impetus for introduction of the latter notation will become apparent in the next section. For this presentation it suffices to assume a single infinite precision map, however there indeed may also be numerical approximation errors, i.e. Ψ^L may be used in place of Ψ for some satisfactory resolution L . The prediction step is completed by using the particle paths to compute sample mean and covariance operator:

$$\begin{aligned} m_{n+1}^{\text{MC}} &= E_M[v_{n+1}] , \\ C_{n+1}^{\text{MC}} &= \text{Cov}_M[v_{n+1}] , \end{aligned}$$

with the unbiased sample moments

$$E_M[v] := \frac{1}{M} \sum_{i=1}^M v(\omega_i) , \quad (2.11)$$

and

$$\text{Cov}_M[u, v] := \frac{M}{M-1} (E_M[u \otimes v] - E_M[u] \otimes E_M[v]) , \quad (2.12)$$

and the shorthand $\text{Cov}_M[u] := \text{Cov}_M[u, u]$. The update step consists of computing (1) auxiliary operators

$$S_{n+1}^{\text{MC}} = H C_{n+1}^{\text{MC}} H^* + \Gamma \text{ and } K_{n+1}^{\text{MC}} = (C_{n+1}^{\text{MC}} H^*) (S_{n+1}^{\text{MC}})^{-1} , \quad (2.13)$$

where H^* is the adjoint of H defined by $\langle a, Hu \rangle_{\mathcal{R}_m} = \langle H^* a, u \rangle_{\mathcal{H}}$ for all $a \in \mathbb{R}^m$ and $u \in \mathcal{H}$, and (2) measurement corrected particle paths for $i = 1, 2, \dots, M$,

$$\begin{aligned} \tilde{y}_{n+1,i} &= y_{n+1} + \eta_{n+1,i} , \\ \hat{v}_{n+1,i} &= (I - K_{n+1}^{\text{MC}} H) v_{n+1,i} + K_{n+1}^{\text{MC}} \tilde{y}_{n+1,i} , \end{aligned}$$

where the sequence $\{\eta_{n+1,i}\}_{i=1}^M$ is i.i.d. with $\eta_{n+1,1} \sim N(0, \Gamma)$. This last procedure may appear somewhat ad-hoc. Indeed it was originally introduced in [5] to correct the statistical error induced in its absence in implementations following the original formulation of the ensemble Kalman filter in

¹Due to the implicit linear and Gaussian assumptions underlying the formulation, one may determine that it is reasonable to summarize the ensemble by its sample mean and covariance and indeed this is often done. In this case, one may construct a Gaussian from the empirical statistics and resample from that.

[12]. It has become known as the perturbed observation implementation. Due to the form of the update, all ensemble members are correlated to one another after the first update. So, even in the linear Gaussian case, the ensemble is no longer Gaussian after the first update. Nonetheless, it has been shown that the limiting ensemble converges to the correct Gaussian in the linear and finite-dimensional case [37, 33], with the rate $\mathcal{O}(N^{-1/2})$ in L^p for Lipschitz functionals with polynomial growth at infinity. Furthermore, it converges with the same rate in the nonlinear but Lipschitz case, i.e. under Assumption 1 [33, 32], to a limiting distribution which will be discussed further in the subsection 2.5. The measurement corrected sample mean and covariance, which need not be computed, would be given by:

$$\begin{aligned}\hat{m}_{n+1}^{\text{MC}} &= E_M[\hat{v}_{n+1}], \\ \hat{C}_{n+1}^{\text{MC}} &= \text{Cov}_M[\hat{v}_{n+1}].\end{aligned}$$

For later computing quantities of interest, we introduce the following notation for the empirical measure of the EnKF ensemble $\{\hat{v}_{n,i}\}_{i=1}^M$.²

$$\hat{\mu}_n^{\text{MC}} = \frac{1}{M} \sum_{i=1}^{M_0} \delta_{\hat{v}_{n,i}}. \quad (2.14)$$

And for any $\varphi : \mathcal{H} \rightarrow \mathbb{R}$, let

$$\hat{\mu}_n^{\text{MC}}(\varphi) := \int \varphi d\mu_n^{\text{ML}} = \frac{1}{M} \sum_{i=1}^M \varphi(\hat{v}_{n,i}).$$

This section is concluded with a comment regarding the required computation of auxiliary operators (2.13). In particular, it will be convenient to introduce index summation notation so that it is assumed that indices which appear twice will be summed over, i.e. $a_k b_k := \sum_k a_k b_k$. Letting $\{e_i\}_{i=1}^m$ be a basis for \mathbb{R}^m , one can write $H = H_{ik} e_i \otimes \phi_k$, where $H_{ik} := \langle e_i, H \phi_k \rangle$, and $C_{n+1}^{\text{MC}} = C_{n+1,kl}^{\text{MC}} \phi_k \otimes \phi_l$, where $C_{n+1,kl}^{\text{MC}} := \langle \phi_k, C_{n+1}^{\text{MC}} \phi_l \rangle$. Then it makes sense to define the intermediate operator

$$R_{n+1}^{\text{MC}} = R_{n+1,ki}^{\text{MC}} \phi_k \otimes e_i, \quad (2.15)$$

where $R_{n+1,ki}^{\text{MC}} = C_{n+1,kl}^{\text{MC}} H_{il}$. The operators of (2.13) can be written in terms of indices as

$$S_{n+1,ij}^{\text{MC}} = H_{il} R_{n+1,lj}^{\text{MC}} + \Gamma_{ij} \text{ and } K_{n+1,ki}^{\text{MC}} = R_{n+1,kg}^{\text{MC}} ((S_{n+1}^{\text{MC}})^{-1})_{gi}, \quad (2.16)$$

where the ranges of the indices $k, l = 1, 2, \dots$ and $i, j, g = 1, 2, \dots, m$ are understood.

²Similar may be done for the predicting distributions, but the updated distributions will be our primary interest.

2.4. Multilevel EnKF. Herein a hierarchy of spaces are introduced $\mathcal{H}_\ell = \text{span}\{\phi_l\}_{l=1}^{N_\ell}$, where $\{N_\ell\}$ is an exponentially increasing sequence of natural numbers further described in Assumption 2. Define $\Phi_\ell = [\phi_1, \dots, \phi_{N_\ell}] : \mathbb{R}^{N_\ell} \rightarrow \mathcal{H}$ and the projection operator $\mathcal{P}_\ell := \Phi_\ell \Phi_\ell^\top$. For $u \in \mathcal{H}$, $u^\ell = \mathcal{P}_\ell u = \sum_{l=1}^{N_\ell} u_l \phi_l \in \mathcal{H}_\ell$, where $u_l = \langle \phi_l, u \rangle$. One has that $\mathcal{H} \supset \dots \supset \mathcal{H}_{\ell+1} \supset \mathcal{H}_\ell \supset \dots \supset \mathcal{H}_0$. MLEnKF computes particle paths on this hierarchy of spaces with a hierarchy of accuracy levels. The case where the accuracy levels are given by refinement of the temporal discretization has already been covered in [20], for finite-dimensional state space. Let v_n^ℓ, \hat{v}_n^ℓ respectively denote the prediction and update of a particle on solution level ℓ at simulation time n . A solution on level ℓ is computed by the numerical integrator $v_{n+1}^\ell = \Psi^\ell(\hat{v}_n^\ell)$. Furthermore, let the increment operator for level ℓ be given by

$$\Delta_\ell v_n := \begin{cases} v_n^0, & \text{if } \ell = 0, \\ v_n^\ell - v_n^{\ell-1}, & \text{else if } \ell > 0. \end{cases}$$

Then the transition from approximation of the distribution of $u_n|Y_n$ to the distribution of $u_{n+1}|Y_{n+1}$ in the MLEnKF framework consists of the predict/update step of generating *pairwise coupled* particle realizations on a set of levels $\ell = 0, 1, \dots, L$. However, it is important to note that here one has correlation between pairs and also between levels due to the update, unlike the standard MLMC in which one has i.i.d. pairs. This point will be very important, and we return to it in the following section.

Similarly to the standard EnKF, the MLEnKF transition is between *multilevel* ensembles $\{(\hat{v}_{n,i}^\ell)_{i=1}^{M_\ell}\}_{\ell=1}^L \mapsto \{(v_{n+1,i}^\ell)_{i=1}^{M_\ell}\}_{\ell=1}^L$. This consists, as for EnKF, of the predict and update steps. In the predict step, particle paths are first computed on a hierarchy of levels. That is, the particle paths are computed one step forward by

$$\begin{aligned} v_{n+1}^{\ell-1}(\omega_{\ell,i}) &= \Psi^{\ell-1}(\hat{v}_n^{\ell-1}(\omega_{\ell,i}), \omega_{\ell,i}), \\ v_{n+1}^\ell(\omega_{\ell,i}) &= \Psi^\ell(\hat{v}_n^\ell(\omega_{\ell,i}), \omega_{\ell,i}), \end{aligned} \tag{2.17}$$

for the levels $\ell = 0, 1, \dots, L$ and level particles $i = 1, 2, \dots, M_\ell$ (where for convenience we introduce the convention that $v^{-1} := 0$). Here the noise in the second argument of the Ψ^ℓ is correlated only within pairs, and are otherwise independent. Thereafter, sample mean and covariance matrices are computed as a sum of sample moments of increments over all levels:

$$\begin{aligned} m_{n+1}^{\text{ML}} &= \sum_{\ell=0}^L E_{M_\ell}[\Delta_\ell v_{n+1}(\omega_{\ell,\cdot})], \\ C_{n+1}^{\text{ML}} &= \sum_{\ell=0}^L \text{Cov}_{M_\ell}[v_{n+1}^\ell(\omega_{\ell,\cdot})] - \text{Cov}_{M_\ell}[v_{n+1}^{\ell-1}(\omega_{\ell,\cdot})], \end{aligned} \tag{2.18}$$

where we recall the sample moment notation (2.11) and (2.12). Define

$$X_{M_\ell} := \frac{1}{\sqrt{M_\ell - 1}} \left([v_{n+1}^\ell(\omega_{\ell,1}), \dots, v_{n+1}^\ell(\omega_{\ell,M_\ell})] - E_{M_\ell}[v_{n+1}^\ell(\omega_{\ell,\cdot})] \mathbf{1}^\top \right), \quad (2.19)$$

where $\mathbf{1}$ is a vector of M_ℓ ones. Then $\text{Cov}_{M_\ell}[v_{n+1}^\ell(\omega_{\ell,\cdot})] = X_{M_\ell} X_{M_\ell}^\top$. The cost of construction is $N_\ell^2 \times M_\ell$, and would therefore be the dominant level ℓ cost. It turns out it is not necessary to construct the full covariance, as will be described below.

Recalling (2.13) and (2.16), it is necessary for the stability of the algorithm that the matrix HR_n^{ML} appearing in the denominator of the gain (where R_n^{ML} is the multilevel version of the operator defined in the Monte Carlo context in equation (2.15)) is positive semi-definite, a condition which is *not* guaranteed for multilevel estimators. This will therefore be *imposed* in the algorithm, similarly to the strategy in the recent work [20]. Let

$$HR_n^{\text{ML}} = \sum_{i=1}^m \lambda_i q_i q_i^\top$$

denote the eigenvalue decomposition of HR_n^{ML} . Notice that the multilevel covariance does not ensure $\min_i(\lambda_i) \not\geq 0$. Define

$$H\tilde{R}_n^{\text{ML}} = \sum_{i=1; \lambda_i \geq 0}^m \lambda_i q_i q_i^\top. \quad (2.20)$$

In the update step the multilevel Kalman gain is defined as follows

$$K_{n+1}^{\text{ML}} = R_{n+1}^{\text{ML}} (S_{n+1}^{\text{ML}})^{-1}, \text{ where } S_{n+1}^{\text{ML}} := H\tilde{R}_{n+1}^{\text{ML}} + \Gamma. \quad (2.21)$$

Next, all particle paths are corrected according to measurements and perturbed observations are added:

$$\begin{aligned} \tilde{y}_{n+1,i}^\ell &= y_{n+1} + \eta_{n+1,i}^\ell \\ \hat{v}_{n+1}^{\ell-1}(\omega_{i,\ell}) &= (I - \mathcal{P}_{\ell-1} K_{n+1}^{\text{ML}} H) v_{n+1}^{\ell-1}(\omega_{i,\ell}) + \mathcal{P}_{\ell-1} K_{n+1}^{\text{ML}} \tilde{y}_{n+1,i}^\ell, \\ \hat{v}_{n+1}^\ell(\omega_{i,\ell}) &= (I - \mathcal{P}_\ell K_{n+1}^{\text{ML}} H) v_{n+1}^\ell(\omega_{i,\ell}) + \mathcal{P}_\ell K_{n+1}^{\text{ML}} \tilde{y}_{n+1,i}^\ell, \end{aligned} \quad (2.22)$$

where the sequence $\{\eta_{n+1,i}^\ell\}_{i=1}^N$ is i.i.d. with $\eta_{n+1,1}^{\{0\}} \sim N(0, \Gamma)$. It is in this step precisely that the pairs all become correlated with one another and the situation becomes significantly more complex than the i.i.d. case. After the first update, this correlation propagates forward through (2.17) to the next observation time via this ensemble. This is the conclusion of the update step of the MLEnKF, and this multilevel ensemble is subsequently propagated forward to the next prediction time via (2.17).

Proposition 2. *Assuming $m \ll N_0$, the cost arising from level ℓ in the construction of the M_ℓ sample updates (2.22) is proportional to $m \times N_\ell \times M_\ell$.*

Proof. Two separate operations are required at each level ℓ . The first arises in the construction of the multilevel gain K_{n+1}^{ML} in (2.21). Now shall become

apparent the impetus for introducing the operator $R_{n+1}^{\text{ML}} = C_{n+1}^{\text{ML}} H^*$ in (2.15). Notice at no point is the full C_{n+1}^{ML} required, but rather only

$$R_{n+1}^{\text{ML}} = \sum_{\ell=0}^L \text{Cov}_{M_\ell}[v_{n+1}^\ell(\omega_{\ell,\cdot}), H v_{n+1}^\ell(\omega_{\ell,\cdot})] - \text{Cov}_{M_\ell}[v_{n+1}^{\ell-1}(\omega_{\ell,\cdot}), H v_{n+1}^{\ell-1}(\omega_{\ell,\cdot})].$$

The level ℓ contribution to this is dominated by the operation $X_{M_\ell}(H X_{M_\ell})^\top$, where X_{M_ℓ} is defined in (2.19). The cost of constructing $H X_{M_\ell} \in \mathbb{R}^{m \times M_\ell}$ is proportional to $m \times N_\ell \times M_\ell$, and so the cost of constructing $X_{M_\ell}(H X_{M_\ell})^\top$ is proportional to $2 \times m \times N_\ell \times M_\ell$. There is also an insignificant one time cost of $\mathcal{O}(m^2 N_L)$ in the construction and inversion of S_{n+1}^{ML} .

The second operation at level ℓ arises from actually computing the update (2.22) using $\mathcal{P}_\ell K_{n+1}^{\text{ML}}$. The cost of obtaining $\mathcal{P}_\ell K_{n+1}^{\text{ML}}$ from K_{n+1}^{ML} is negligible, so it is clear that each sample incurs a cost $m \times N_\ell$. \square

The following notation denotes the empirical measure of the multilevel ensemble $\{(\hat{v}_{n,i}^\ell)_{i=1}^{M_\ell}\}_{\ell=1}^L$:

$$\hat{\mu}_n^{\text{ML}} = \frac{1}{M_0} \sum_{i=1}^{M_0} \delta_{\hat{v}_n^0(\omega_{i,0})} + \sum_{\ell=1}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} (\delta_{\hat{v}_n^\ell(\omega_{i,\ell})} - \delta_{\hat{v}_n^{\ell-1}(\omega_{i,\ell})}), \quad (2.23)$$

and for any $\varphi : \mathcal{H} \rightarrow \mathbb{R}$, let

$$\hat{\mu}_n^{\text{ML}}(\varphi) := \int \varphi d\mu_n^{\text{ML}} = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} \varphi(\hat{v}_n^\ell(\omega_{i,\ell})) - \varphi(\hat{v}_n^{\ell-1}(\omega_{i,\ell})).$$

2.5. Nonlinear Kalman filtering. It will be useful to introduce the limiting process, in the case of nonlinear non-Gaussian forward model (2.1). The following process defines the MFEnKF [32]:

$$\text{Prediction} \begin{cases} \bar{v}_{n+1} &= \Psi(\hat{v}_n), \\ \bar{m}_{n+1} &= \mathbb{E}[\bar{v}_{n+1}], \\ \bar{C}_{n+1} &= \mathbb{E}[(\bar{v}_{n+1} - \bar{m}_{n+1}) \otimes (\bar{v}_{n+1} - \bar{m}_{n+1})] \end{cases} \quad (2.24)$$

$$\text{Update} \begin{cases} \bar{S}_{n+1} &= (H \bar{C}_{n+1}) H^* + \Gamma \\ \bar{K}_{n+1} &= (\bar{C}_{n+1} H^*) \bar{S}_{n+1}^{-1} \\ \tilde{y}_{n+1} &= y_{n+1} + \eta_{n+1} \\ \hat{v}_{n+1} &= (I - \bar{K}_{n+1} H) \bar{v}_{n+1} + \bar{K}_{n+1} \tilde{y}_{n+1}. \end{cases} \quad (2.25)$$

Here η_n are i.i.d. draws from $N(0, \Gamma)$. It is easy to see that in the linear Gaussian case the mean and variance of the above process correspond to the mean and variance of the filtering distribution [30]. Moreover, it was shown in [37, 33] that for finite-dimensional state-space the single level EnKF converges to the Kalman filtering distribution with the standard rate $\mathcal{O}(M^{-1/2})$ in this case. It was furthermore shown in [33] and [32] that for nonlinear Gaussian state-space models and fully non-Gaussian models (2.1), respectively, the EnKF converges to the above process with the same rate as long as the models satisfy a Lipschitz criterion as in Assumption 1. The work of

[20] illustrated that the MLEnKF converges as well, and with an asymptotic cost-to- ε which is strictly smaller than its single level EnKF counterpart. The work of [29] extended convergence results to infinite-dimensional state-space for square root filters. In this work, the aim is to prove convergence of the MLEnKF for infinite-dimensional state-space, with the same favorable asymptotic cost-to- ε performance.

The following fact will be necessary in the subsequent section.

Proposition 3. *Given Assumption 1 on Ψ , the MFEnKF process (2.24)–(2.25) satisfies $\bar{v}_n, \hat{v}_n \in L^p(\Omega; \mathcal{H})$ for all $n \in \mathbb{N}$.*

Proof. Clearly it holds for time $n = 0$. Given $\hat{v}_n \in L^p(\Omega; \mathcal{H})$, Assumption 1 (ii) guarantees $\bar{v}_{n+1} \in L^p(\Omega; \mathcal{H})$. By Proposition 1, $\bar{C}_{n+1} \in \mathcal{H} \otimes \mathcal{H}$. Since $H\bar{C}_{n+1}H^* \geq 0$ and $\Gamma > 0$, it is clear that $\bar{S}_{n+1} > 0$, which implies $\|H^*\bar{S}_{n+1}^{-1}\|_{\mathcal{H} \otimes \mathcal{R}_m} < \infty$. Hence, $\bar{K}_{n+1} \in \mathcal{H} \otimes \mathcal{R}_m$. Therefore it is clear that $\hat{v}_{n+1} \in L^p(\Omega; \mathcal{H})$. \square

3. THEORETICAL RESULTS

The approximation error and computational cost of approximating the true filtering distribution by MLEnKF when given a sequence of observations y_1, y_2, \dots, y_n will be studied in this section. Before stating the main approximation theorem, it will be useful to present the basic assumptions that will be used throughout and the corresponding standard MLMC approximation results for i.i.d. samples, as well as a slight variant which will be useful in what follows.

Definition 1. *A function $\varphi : \mathcal{H} \rightarrow \mathbb{R}$ is said to be globally Lipschitz continuous provided there exist a positive scalar $C_\varphi < \infty$ such that for all $u, v \in \mathcal{H}$*

$$|\varphi(u) - \varphi(v)| \leq C_\varphi \|u - v\|_{\mathcal{H}}. \quad (3.1)$$

Assumption 2. *Consider the hidden Markov model defined by (2.1) and (2.2) with initial data $u_0 \in L^p(\Omega; \mathcal{H})$ for all $p \geq 2$ and assume that the sequence of resolution dimensions $\{N_\ell\}$ fulfils the exponential growth constraint $N_\ell \asymp \kappa^\ell$, for some $\kappa > 1$. Let Ψ^ℓ denote a numerical solver with a resolution parameter $h_\ell \asymp N_\ell^{-1/d}$. This will define the hierarchy of solution operators in Section 2, which are assumed to satisfy Assumption 1. For a given set of constants $\beta, \gamma > 0$, assume the following conditions are fulfilled for all $\ell \geq 0$ and $u, v \in L^p(\Omega; \mathcal{H})$ for all $p \geq 2$:*

- (i) $\|\Psi^\ell(u) - \Psi(u)\|_{L^p(\Omega; \mathcal{H})} \lesssim h_\ell^{\beta/2}$, for all $p \geq 2$,
- (ii) $\|(I - \mathcal{P}_\ell)u_0\|_{L^p(\Omega; \mathcal{H})} \lesssim h_\ell^{\beta/2}$, for all $p \geq 2$,
- (iii) $\mathcal{C}_\ell \lesssim h_\ell^{-d\gamma}$, where \mathcal{C}_ℓ denotes the computational cost associated to level ℓ (and d is the spatiotemporal dimension of the continuum which is being approximated)³.

³This can be made much more general, but the objective here is simplicity of exposition

Assumption 2 is given in a bare-minimum form, which we believe will be easier to verify when applying the method to particular problems. The next corollary states direct consequences of the above assumption, which will be useful for proving properties of the MLEnKF method.

Proposition 4. *Suppose Assumption 2 holds and $\Psi^\ell = \mathcal{P}_\ell \Psi$. Then for all $\ell \in \mathbb{N}$, $u, v \in L^p(\Omega; \mathcal{H})$ for all $p \geq 2$, and globally Lipschitz continuous observables φ :*

- (i) $\|\Psi^\ell(v) - \Psi^{\ell-1}(v)\|_{L^p(\Omega; \mathcal{H})} \lesssim h_\ell^{\beta/2}$, for all $p \geq 2$,
- (ii) $|\mathbb{E}[\varphi(\Psi^\ell(u)) - \varphi(\Psi(v))]| \lesssim \|u - v\|_{L^p(\Omega; \mathcal{H})} + h_\ell^{\beta/2}$, for all $p \geq 2$,
- (iii) $\|(I - \mathcal{P}_\ell)\bar{C}_n\|_{\mathcal{H} \otimes \mathcal{H}} \lesssim h_\ell^{\beta/2}$.

Proof. Property (i) follows from Assumption 2(i) and Minkowski's inequality. Property (ii) follows from Definition 1, followed by Minkowski's inequality, Assumption 1(i), and Assumption 2(i). For property (iii), recall Proposition 3, and without loss of generality assume $\mathbb{E}[\bar{v}_n] = 0$ (for simplicity of the argument to follow). Now observe

$$\|(I - \mathcal{P}_\ell)\bar{C}_n\|_{\mathcal{H} \otimes \mathcal{H}} = \|\mathbb{E}[(I - \mathcal{P}_\ell)\bar{v}_n \otimes \bar{v}_n]\|_{\mathcal{H} \otimes \mathcal{H}} \leq \|(I - \mathcal{P}_\ell)\bar{v}_n\|_2 \|\bar{v}_n\|_2,$$

where the inequality is a result of Jensen's inequality, the definition (2.6), and Hölder's inequality. Notice that $(I - \mathcal{P}_\ell)\bar{v}_n = (I - \mathcal{P}_\ell)\Psi(\hat{v}_{n-1})$. Since it is assumed that $\Psi^\ell = \mathcal{P}_\ell \Psi$, the claim follows from Assumption 2(i) again. \square

Remark 1. *It will be assumed that the computational cost of the forward simulation, $\text{Cost}(\Psi^\ell) = \mathcal{O}(h_\ell^{-d\gamma})$ is at least linear in N_ℓ , i.e., that $\gamma \geq 1$, and that $m \ll N_0$. Therefore, in view of Proposition 2, the total cost is dominated by $\mathcal{C}_\ell = \mathcal{O}(h_\ell^{-d\gamma})$. It is important to observe that in the big data case $m \geq N_0$, the algorithm will need to be modified to be efficient in the non-asymptotic regime when the accuracy constraint ε , relatively speaking, is large. For larger values of m , smaller ε regimes will be affected.*

We will now state the main theorem of this paper. It gives an upper bound for the computational cost of achieving a sought accuracy in $L^p(\Omega; \mathcal{H})$ -norm when using the MLEnKF method to approximate the expectation of an observable. The theorem may be considered an extension to spatially extended models of the earlier work [20].

Theorem 1 (MLEnKF accuracy vs. cost). *Consider a globally Lipschitz continuous observable function $\varphi : \mathcal{H} \rightarrow \mathbb{R}$, and suppose Assumptions 1 and 2 hold. For a given $\varepsilon > 0$, let L and $\{M_\ell\}_{\ell=0}^L$ be defined under the constraints $L = \lceil 2 \log_\kappa(\varepsilon^{-1})/\beta \rceil$ and*

$$M_\ell \asymp \begin{cases} h_\ell^{(\beta+d\gamma)/2} h_L^{-\beta}, & \text{if } \beta > d\gamma, \\ h_\ell^{(\beta+d\gamma)/2} L^2 h_L^{-\beta}, & \text{if } \beta = d\gamma, \\ h_\ell^{(\beta+d\gamma)/2} h_L^{-(\beta+d\gamma)/2}, & \text{if } \beta < d\gamma. \end{cases} \quad (3.2)$$

Then, for any $p \geq 2$ and $n \in \mathbb{N}$,

$$\|\hat{\mu}_n^{\text{ML}}(\varphi) - \hat{\mu}_n(\varphi)\|_{L^p(\Omega; \mathcal{H})} \lesssim |\log(\varepsilon)|^n \varepsilon, \quad (3.3)$$

where $\hat{\mu}_n^{\text{ML}}$ denotes the multilevel empirical measure defined in (2.23) whose particle evolution is given by the multilevel predict (2.17) and update (2.22) formulae, approximating the time n mean-field EnKF distribution $\hat{\mu}_n$ (the filtering distribution $\hat{\mu}_n = N(\hat{m}_n, \hat{C}_n)$ in the linear Gaussian case).

The computational cost of the MLEnKF estimator over the time sequence is bounded by

$$\text{Cost (MLEnKF)} \lesssim \begin{cases} \varepsilon^{-2}, & \text{if } \beta > d\gamma, \\ \varepsilon^{-2} |\log(\varepsilon)|^3, & \text{if } \beta = d\gamma, \\ \varepsilon^{-2d\gamma/\beta}, & \text{if } \beta < d\gamma. \end{cases} \quad (3.4)$$

The proof follows very closely that of [20, Theorem 3.2], except here it is extended to the Hilbert space setting with approximation of spatially extended models.

Following [20] and [33, 32, 37], introduce the mean-field limiting multilevel ensemble $\{(\bar{v}_{n,i}^\ell)_{i=1}^{M_\ell}\}_{\ell=1}^L$, which evolves according to the same equations with the same realizations of noise except the covariance \bar{C}_n , and hence the Kalman gain \bar{K}_n , are given by limiting formulae in (2.24) and (2.25). An ensemble member \bar{v}^ℓ corresponds to a solution of the above system with $\bar{v}_{n+1}^\ell = \Psi^\ell(\hat{v}_n^\ell)$ replacing the first equation and the equation

$$\hat{v}_{n+1}^\ell = (I - \mathcal{P}_\ell \bar{K}_{n+1} H) \bar{v}_{n+1}^\ell + \mathcal{P}_\ell \bar{K}_{n+1} \tilde{y}_{n+1}^\ell$$

replacing the last equation. The sample $\bar{v}^\ell(\omega_{\ell,i})$ is a single realization of this system above with the same noise realization $\omega_{\ell,i}$ as the sample $v^\ell(\omega_{\ell,i})$ from MLEnKF, including the perturbed observation. Note that the processes \bar{v}^ℓ , \hat{v}^ℓ are bounded in $L^p(\Omega; \mathcal{H})$ as well by similar arguments to Proposition 3.

Let us first recall that the multilevel Kalman gain is defined by

$$K_n^{\text{ML}} = R_n^{\text{ML}} (H \tilde{R}_n^{\text{ML}} + \Gamma)^{-1}$$

where

$$H \tilde{R}_n^{\text{ML}} = \sum_{i=1; \lambda_i > 0}^m \lambda_i q_i q_i^\top, \quad (3.5)$$

for eigenpairs $\{\lambda_i, q_i\}$ of $H R_n^{\text{ML}}$. The following micro-lemma will be necessary to control the error in the gain.

Lemma 1 (Multilevel covariance approximation error). *Let \tilde{R}_n^{ML} be given by (3.5). Then there is a $0 < c < \infty$ such that*

$$\left\| H(\tilde{R}_n^{\text{ML}} - R_n^{\text{ML}}) \right\|_{\mathcal{R}_m \otimes \mathcal{R}_m} \leq c \|C_n^{\text{ML}} - \bar{C}_n\|_{\mathcal{H} \otimes \mathcal{H}}. \quad (3.6)$$

Proof. Notice that, by equivalence of 2 and HS norms over \mathbb{R}^m , there exists a $0 < \tilde{c} < \infty$ such that

$$\left\| H(\tilde{R}_n^{\text{ML}} - R_n^{\text{ML}}) \right\|_{\mathcal{R}_m \otimes \mathcal{R}_m} \leq \tilde{c} \max_{\{j; \lambda_j < 0\}} \{|\lambda_j|\}. \quad (3.7)$$

Denote the associated eigenvector by q_{\max} (normalized to $\|q_{\max}\| = 1$). Notice that for any $A = A^\top \in \mathbb{R}^{m \times m}$ we can define

$$\|A\| := \sup_q \frac{|q^\top A q|}{\|q\|^2} = \max_i |\lambda_i|,$$

where λ_i are the eigenvalues of A .

Since $\bar{C}_n \geq 0$, one has that

$$\begin{aligned} \left| q_{\max}^\top H(C_n^{\text{ML}} - \bar{C}_n) H^* q_{\max} \right| &= q_{\max}^\top H \bar{C}_n H^* q_{\max} - q_{\max}^\top H C_n^{\text{ML}} H^* q_{\max} \\ &\geq \frac{1}{\tilde{c}} \left\| H(\tilde{R}_n^{\text{ML}} - R_n^{\text{ML}}) \right\|_{\mathcal{R}_m \otimes \mathcal{R}_m}. \end{aligned}$$

The fact that for self-adjoint $Q : \mathcal{H} \rightarrow \mathcal{H}$ one has $\|HQH^*\|_{\mathcal{R}_m \otimes \mathcal{R}_m} \leq \|H\|_{\mathcal{R}_m \otimes \mathcal{H}}^2 \|Q\|_{\mathcal{H} \otimes \mathcal{H}}$ concludes the proof. \square

The next step is to bound the Kalman gain error in terms of the covariance error.

Lemma 2 (Kalman gain error). *There is a constant $\tilde{c}_n < \infty$, depending on $\|H\|_{\mathcal{R}_m \otimes \mathcal{H}}$, γ_{\min} , and $\|\bar{K}_n H\|_{\mathcal{H} \otimes \mathcal{H}}$ such that*

$$\|K_n^{\text{ML}} - \bar{K}_n\|_{\mathcal{H} \otimes \mathcal{R}_m} \leq \tilde{c}_n \|C_n^{\text{ML}} - \bar{C}_n\|_{\mathcal{H} \otimes \mathcal{H}}. \quad (3.8)$$

Proof. It is shown in Lemma 3.4 of [20] that

$$\bar{K}_n - K_n^{\text{ML}} = \bar{K}_n H (\tilde{R}_n^{\text{ML}} - R_n) (H \tilde{R}_n^{\text{ML}} + \Gamma)^{-1} \quad (3.9)$$

$$+ ((\bar{C}_n - C_n^{\text{ML}}) H^*) (H \tilde{R}_n^{\text{ML}} + \Gamma)^{-1}. \quad (3.10)$$

Note that $x^\top (\Gamma + B)x \geq x^\top \Gamma x \geq \gamma_{\min}$ for all $x \in \mathbb{R}^m$ whenever $B = B^\top \geq 0$, and this implies that $\left\| (H \tilde{R}_n^{\text{ML}} H^* + \Gamma)^{-1} \right\|_{\mathcal{R}_m \otimes \mathcal{R}_m} \leq 1/\gamma_{\min}$ where $\gamma_{\min} > 0$ is the smallest eigenvalue of Γ . It follows by (3.6) that

$$\|\bar{K}_n - K_n^{\text{ML}}\|_{\mathcal{H} \otimes \mathcal{R}_m} \leq \frac{1 + 2 \|\bar{K}_n H\|_{\mathcal{H} \otimes \mathcal{H}}}{\gamma_{\min}} \|H\|_{\mathcal{R}_m \otimes \mathcal{H}} \|\bar{C}_n - C_n^{\text{ML}}\|_{\mathcal{H} \otimes \mathcal{H}}. \quad (3.11)$$

\square

Theorem 2. *Suppose Assumptions 1 and 2 hold and for any $\varepsilon > 0$, let L and $\{M_\ell\}_{\ell=0}^L$ be defined as in Theorem 1. Then the following inequality holds for any $p \geq 2$ and $n \in \mathbb{N}$,*

$$\|C_n^{\text{ML}} - \bar{C}_n\|_{L^p(\Omega; \mathcal{H} \otimes \mathcal{H})} \lesssim \varepsilon + \|C_n^{\text{ML}} - \bar{C}_n^{\text{ML}}\|_{L^p(\Omega; \mathcal{H} \otimes \mathcal{H})}. \quad (3.12)$$

Proof. Let \bar{C}_n^L denote the predicting covariance of the final L^{th} level limiting system at time n , in the sense that the forward map above is replaced by Ψ_L , but the gain comes from the continuum mean-field limiting system. Furthermore, let \bar{C}_n^{ML} denote the covariance associated to the multilevel ensemble $\{(\bar{v}_{n,i}^\ell)_{i=1}^{M_\ell}\}_{\ell=1}^L$. Minkowski's inequality is used to split

$$\|C_n^{ML} - \bar{C}_n\|_p \leq \|\bar{C}_n^L - \bar{C}_n\|_p + \|\bar{C}_n^{ML} - \bar{C}_n^L\|_p + \|C_n^{ML} - \bar{C}_n^{ML}\|_p, \quad (3.13)$$

and each term will be dealt with in turn, in the following three lemmas. The proof of the theorem is concluded after Lemmas 3 and 4 which bound the first two terms, respectively. \square

Lemma 3. *Suppose Assumptions 1 and 2 hold and for any $\varepsilon > 0$, let L be defined as in Theorem 1. Then the following inequalities hold for any $n \in \mathbb{N}$ and $p \geq 2$,*

$$\|\bar{C}_n^L - \bar{C}_n\|_{\mathcal{H} \otimes \mathcal{H}} \lesssim \varepsilon, \quad (3.14)$$

$$\max(\|\bar{v}_n^L - \bar{v}_n\|_{L^p(\Omega; \mathcal{H})}, \|\hat{v}_n^L - \hat{v}_n\|_{L^p(\Omega; \mathcal{H})}) \lesssim \varepsilon, \quad (3.15)$$

and

$$\max\left(\|\bar{v}_n^\ell - \bar{v}_n^{\ell-1}\|_{L^p(\Omega; \mathcal{H})}, \|\hat{v}_n^\ell - \hat{v}_n^{\ell-1}\|_{L^p(\Omega; \mathcal{H})}\right) \lesssim h_\ell^{\beta/2}, \quad \forall \ell \in \mathbb{N}. \quad (3.16)$$

Proof. The initial data for the respective mean-field methods is given by \hat{v}_0 and $\hat{v}_0^L =: \mathcal{P}_L \hat{v}_0$. Assumption 2(ii) implies that

$$\|\hat{v}_0 - \hat{v}_0^L\|_p \lesssim h_L^{\beta/2} \lesssim \varepsilon.$$

By Assumptions 1(i) and 2(i),

$$\|\bar{v}_n - \bar{v}_n^L\|_p \lesssim \|\hat{v}_{n-1} - \hat{v}_{n-1}^L\|_p + h_L^{\beta/2},$$

and by Proposition 4(iii),

$$\begin{aligned} \|\hat{v}_n - \hat{v}_n^L\|_p &\leq \|I - \bar{K}_n H\|_{\mathcal{H} \otimes \mathcal{H}} \|\bar{v}_n^L - \bar{v}_n\|_p + \|(I - \mathcal{P}_L) \bar{K}_n (H \bar{v}_n^L + y_n)\|_p \\ &\leq c (\|\bar{v}_n^L - \bar{v}_n\|_p + \|(I - \mathcal{P}_L) \bar{C}_n\|_{\mathcal{H} \times \mathcal{H}}) \\ &\lesssim \|\bar{v}_n^L - \bar{v}_n\|_p + \varepsilon, \end{aligned}$$

where $\bar{S}_n := (H \bar{C}_n H^* + \Gamma)$. Inequality (3.15) consequently holds by induction. Furthermore,

$$\begin{aligned} &\|\bar{C}_n^L - \bar{C}_n\|_{\mathcal{H} \otimes \mathcal{H}} \\ &= \|\mathbb{E}[(\bar{v}_n^L - \mathbb{E}[\bar{v}_n^L]) \otimes (\bar{v}_n^L - \mathbb{E}[\bar{v}_n^L]) - (\bar{v}_n - \mathbb{E}[\bar{v}_n]) \otimes (\bar{v}_n - \mathbb{E}[\bar{v}_n])]\|_{\mathcal{H} \otimes \mathcal{H}} \\ &\leq \|(\bar{v}_n^L - \mathbb{E}[\bar{v}_n^L]) \otimes (\bar{v}_n^L - \mathbb{E}[\bar{v}_n^L]) - (\bar{v}_n - \mathbb{E}[\bar{v}_n]) \otimes (\bar{v}_n - \mathbb{E}[\bar{v}_n])\|_1 \\ &\leq (\|\bar{v}_n^L - \mathbb{E}[\bar{v}_n^L]\|_2 + \|\bar{v}_n - \mathbb{E}[\bar{v}_n]\|_2) \|\bar{v}_n^L - \bar{v}_n\|_2 \\ &\lesssim \varepsilon. \end{aligned}$$

An analogous argument may be used to bound the second term of inequality (3.14).

To prove inequality (3.16), note first that due to the matching initial data, the inequality holds trivially for the update at $n = 0$. By Assumption 1(i), Proposition 4(i), and Minkowski's inequality,

$$\begin{aligned} \left\| \bar{v}_n^\ell - \bar{v}_n^{\ell-1} \right\|_p &\leq \left\| \Psi^\ell(\hat{v}_{n-1}^\ell) - \Psi^{\ell-1}(\hat{v}_{n-1}^\ell) \right\|_p + \left\| \Psi^{\ell-1}(\hat{v}_{n-1}^\ell) - \Psi^{\ell-1}(\hat{v}_{n-1}^{\ell-1}) \right\|_p \\ &\lesssim \left\| \hat{v}_{n-1}^\ell - \hat{v}_{n-1}^{\ell-1} \right\|_p + h_\ell^{\beta/2}, \end{aligned}$$

and by Proposition 4(iii),

$$\begin{aligned} \left\| \hat{v}_n^\ell - \hat{v}_n^{\ell-1} \right\|_p &\leq \left\| I - \mathcal{P}_\ell \bar{K}_n H \right\|_{\mathcal{H} \otimes \mathcal{H}} \left\| \bar{v}_n^\ell - \bar{v}_n^{\ell-1} \right\|_p + \left\| (\mathcal{P}_\ell - \mathcal{P}_{\ell-1}) \bar{K}_n H \bar{v}_n^{\ell-1} \right\|_p \\ &\lesssim \left\| \bar{v}_n^\ell - \bar{v}_n^{\ell-1} \right\|_p + (\left\| (I - \mathcal{P}_\ell) \bar{C}_n \right\|_{\mathcal{H} \otimes \mathcal{H}} + \left\| (I - \mathcal{P}_{\ell-1}) \bar{C}_n \right\|_{\mathcal{H} \otimes \mathcal{H}}) \left\| \bar{v}_n^{\ell-1} \right\|_p \\ &\lesssim \left\| \bar{v}_n^\ell - \bar{v}_n^{\ell-1} \right\|_p + h_\ell^{-\beta/2}. \end{aligned}$$

Inequality (3.16) holds by induction. \square

Next we derive a bound for $\|\bar{C}_n^{\text{ML}} - \bar{C}_n^L\|_p$, where we will make use of the following representation of the finite resolution mean-field covariance

$$\bar{C}_n^L = \sum_{\ell=0}^L \text{Cov}[\bar{v}_n^\ell] - \text{Cov}[\bar{v}_n^{\ell-1}].$$

We also recall that \bar{C}_n^{ML} denotes the mean-field MLEnKF sample covariance defined by

$$\bar{C}_n^{\text{ML}} = \sum_{\ell=0}^L \text{Cov}_{M_\ell}[\bar{v}_n^\ell] - \text{Cov}_{M_\ell}[\bar{v}_n^{\ell-1}]. \quad (3.17)$$

Lemma 4 (Multilevel i.i.d. sample covariance error). *Suppose Assumptions 1 and 2 hold and for any $\varepsilon > 0$, let L and $\{M_\ell\}_{\ell=0}^L$ be defined as in Theorem 1. Then the following inequality holds for any $n \in \mathbb{N}$ and $p \geq 2$,*

$$\|\bar{C}_n^{\text{ML}} - \bar{C}_n^L\|_{L^p(\Omega; \mathcal{H} \otimes \mathcal{H})} \lesssim \varepsilon. \quad (3.18)$$

Proof. Recall that (3.17) is unbiased, $\mathbb{E}[\bar{C}_n^{\text{ML}}] = \bar{C}_n^L$, so

$$\|\bar{C}_n^{\text{ML}} - \bar{C}_n^L\|_p = \|\bar{C}_n^{\text{ML}} - \mathbb{E}[\bar{C}_n^{\text{ML}}]\|_p. \quad (3.19)$$

For a random field $Y : \Omega \rightarrow \mathcal{H}$ we introduce the shorthand $\underline{Y} := Y - \mathbb{E}[Y]$. By equation (3.17),

$$\begin{aligned} \|\bar{C}_n^{\text{ML}} - \mathbb{E}[\bar{C}_n^{\text{ML}}]\|_p &= \left\| \sum_{\ell=0}^L \left(\text{Cov}_{M_\ell}[\bar{v}_n^\ell] - \text{Cov}_{M_\ell}[\bar{v}_n^{\ell-1}] \right) \right\|_p \\ &\leq \sum_{\ell=0}^L \left\| \text{Cov}_{M_\ell}[\bar{v}_n^\ell] - \text{Cov}_{M_\ell}[\bar{v}_n^{\ell-1}] \right\|_p \\ &\leq \sum_{\ell=0}^L \left(\left\| \text{Cov}_{M_\ell}[\bar{v}_n^\ell, \Delta_\ell \bar{v}_n] \right\|_p + \left\| \text{Cov}_{M_\ell}[\Delta_\ell \bar{v}_n, \bar{v}_n^{\ell-1}] \right\|_p \right), \end{aligned} \quad (3.20)$$

where we recall that $\Delta_\ell \bar{v}_n = \bar{v}_n^\ell - \bar{v}_n^{\ell-1}$. We have

$$\text{Cov}_{M_\ell}[\bar{v}_n^\ell, \Delta_\ell \bar{v}_n] = \text{Cov}_{M_\ell}[\bar{v}_n^\ell, \Delta_\ell \bar{v}_n] - \text{Cov}[\bar{v}_n^\ell, \Delta_\ell \bar{v}_n],$$

and similarly for the other term. By Lemmas 3 and 8,

$$\begin{aligned} \|\bar{C}_n^{\text{ML}} - \mathbb{E}[\bar{C}_n^{\text{ML}}]\|_p &\leq 2 \sum_{\ell=0}^L \frac{c}{\sqrt{M_\ell}} (\|\bar{v}_n^\ell\|_{2p} + \|\bar{v}_n^{\ell-1}\|_{2p}) \|\Delta_\ell \bar{v}_n\|_{2p} \\ &\lesssim \sum_{\ell=0}^L \frac{1}{\sqrt{M_\ell}} \|\Delta_\ell \bar{v}_n\|_{2p} \lesssim \sum_{\ell=0}^L M_\ell^{-1/2} h_\ell^{\beta/2} \lesssim \varepsilon. \end{aligned}$$

□

The previous two lemmas complete the proof of Theorem 2. We now turn to bounding the latter term of the right-hand side of inequality (3.12), the difference between multilevel ensemble covariances.

Lemma 5. *Suppose Assumptions 1 and 2 hold and for any $\varepsilon > 0$, let L and $\{M_\ell\}_{\ell=0}^L$ be defined as in Theorem 1. Then, for any $p \geq 2$ and $n \in \mathbb{N}$,*

$$\|C_n^{\text{ML}} - \bar{C}_n^{\text{ML}}\|_{L^p(\Omega; \mathcal{H} \otimes \mathcal{H})} \leq 4 \sum_{\ell=0}^L \|v_n^\ell - \bar{v}_n^\ell\|_{L^{2p}(\Omega, \mathcal{H})} (\|v_n^\ell\|_{L^{2p}(\Omega, \mathcal{H})} + \|\bar{v}_n^\ell\|_{L^{2p}(\Omega, \mathcal{H})}). \quad (3.21)$$

Proof. From the definitions of the sample covariance (2.12) and multilevel sample covariance (2.18), one obtains the bounds

$$\begin{aligned} \|C_n^{\text{ML}} - \bar{C}_n^{\text{ML}}\|_p &\leq \sum_{\ell=0}^L \left(\left\| \text{Cov}_{M_\ell}[v_n^\ell] - \text{Cov}_{M_\ell}[\bar{v}_n^\ell] \right\|_p \right. \\ &\quad \left. + \left\| \text{Cov}_{M_\ell}[v_n^{\ell-1}] - \text{Cov}_{M_\ell}[\bar{v}_n^{\ell-1}] \right\|_p \right), \end{aligned}$$

and

$$\begin{aligned} \left\| \text{Cov}_{M_\ell}[v_n^\ell] - \text{Cov}_{M_\ell}[\bar{v}_n^\ell] \right\|_p &\leq \left\| E_{M_\ell}[v_n^\ell \otimes v_n^\ell] - E_{M_\ell}[\bar{v}_n^\ell \otimes \bar{v}_n^\ell] \right\|_p \\ &\quad + \left\| E_{M_\ell}[v_n^\ell] \otimes E_{M_\ell}[v_n^\ell] - E_{M_\ell}[\bar{v}_n^\ell] \otimes E_{M_\ell}[\bar{v}_n^\ell] \right\|_p \\ &=: I_1 + I_2. \end{aligned}$$

The bilinearity of the sample covariance yields that

$$I_1 \leq \left\| E_{M_\ell}[(v_n^\ell - \bar{v}_n^\ell) \otimes v_n^\ell] \right\|_p + \left\| E_{M_\ell}[\bar{v}_n^\ell \otimes (v_n^\ell - \bar{v}_n^\ell)] \right\|_p \quad (3.22)$$

and

$$I_2 \leq \left\| E_{M_\ell}[(v_n^\ell - \bar{v}_n^\ell)] \otimes E_{M_\ell}[v_n^\ell] \right\|_p + \left\| E_{M_\ell}[\bar{v}_n^\ell] \otimes E_{M_\ell}[(v_n^\ell - \bar{v}_n^\ell)] \right\|_p.$$

For bounding I_1 we use Jensen's and Hölder's inequalities:

$$\begin{aligned} \left\| E_{M_\ell}[(v_n^\ell - \bar{v}_n^\ell) \otimes v_n^\ell] \right\|_p^p &= \mathbb{E} \left[\left\| E_{M_\ell}[(v_n^\ell - \bar{v}_n^\ell) \otimes v_n^\ell] \right\|_{\mathcal{H} \otimes \mathcal{H}}^p \right] \\ &\leq \mathbb{E} \left[E_{M_\ell} \left[\left\| v_n^\ell - \bar{v}_n^\ell \right\|_{\mathcal{H}}^p \left\| v_n^\ell \right\|_{\mathcal{H}}^p \right] \right] \\ &= \mathbb{E} \left[\left\| v_n^\ell - \bar{v}_n^\ell \right\|_{\mathcal{H}}^p \left\| v_n^\ell \right\|_{\mathcal{H}}^p \right] \\ &\leq \left\| v_n^\ell - \bar{v}_n^\ell \right\|_{2p}^p \left\| v_n^\ell \right\|_{2p}^p. \end{aligned}$$

The second summand of inequality (3.22) is bounded similarly, and we obtain

$$I_1 \leq \left\| v_n^\ell - \bar{v}_n^\ell \right\|_{2p} \left(\left\| v_n^\ell \right\|_{2p} + \left\| \bar{v}_n^\ell \right\|_{2p} \right).$$

The I_2 term can also be bounded with similar steps as in the preceding argument so that also

$$I_2 \leq \left\| v_n^\ell - \bar{v}_n^\ell \right\|_{2p} \left(\left\| v_n^\ell \right\|_{2p} + \left\| \bar{v}_n^\ell \right\|_{2p} \right).$$

The proof is finished by summing the contributions of I_1 and I_2 over all levels. \square

It has just been shown that the second term of (3.12) is “close in the predicting ensembles”. Therefore, the error level of the first term will carry over between observation times by induction. This is made rigorous by the next lemma.

Lemma 6 (Distance between ensembles.). *Suppose Assumptions 1 and 2 hold and for any $\varepsilon > 0$, let L and $\{M_\ell\}_{\ell=0}^L$ be defined as in Theorem 1. Then the following inequality holds for any $n \in \mathbb{N}$ and $p \geq 2$,*

$$\sum_{\ell=0}^L \left\| \hat{v}_n^\ell - \hat{\bar{v}}_n^\ell \right\|_{L^p(\Omega; \mathcal{H})} \lesssim |\log(\varepsilon)|^n \varepsilon. \quad (3.23)$$

Proof. We use an induction argument. Notice first of all that by definition,

$$\|v_0^\ell - \bar{v}_0^\ell\|_p = 0$$

Assume that for $p \geq 2$,

$$\sum_{\ell=0}^L \left\| \hat{v}_{n-1}^\ell - \bar{v}_{n-1}^\ell \right\|_p \lesssim |\log(\varepsilon)|^{n-1} \varepsilon. \quad (3.24)$$

By Assumption 1(i), the following inequality holds for the prediction ensemble:

$$\sum_{\ell=0}^L \|v_n^\ell - \bar{v}_n^\ell\|_p \leq \sum_{\ell=0}^L c_\Psi \|\hat{v}_{n-1}^\ell - \bar{v}_{n-1}^\ell\|_p \lesssim |\log(\varepsilon)|^{n-1} \varepsilon. \quad (3.25)$$

Furthermore, by Lemma 2,

$$\begin{aligned} \left\| \hat{v}_n^\ell - \bar{v}_n^\ell \right\|_{\mathcal{H}} &\leq \left\| v_n^\ell - \bar{v}_n^\ell \right\|_{\mathcal{H}} \\ &\quad + \tilde{c}_n \|C_n^{\text{ML}} - C_n\|_{\mathcal{H} \otimes \mathcal{H}} \left(\left\| v_n^\ell - \bar{v}_n^\ell \right\|_{\mathcal{H}} + \left\| y_n^\ell - \bar{v}_n^\ell \right\|_{\mathcal{H}} \right), \end{aligned} \quad (3.26)$$

for all $\ell = 0, \dots, L$. Hölder's inequality then implies

$$\begin{aligned} \|\hat{v}_n^\ell - \bar{v}_n^\ell\|_p &\leq \|v_n^\ell - \bar{v}_n^\ell\|_p \\ &\quad + \tilde{c}_n \|C_n^{\text{ML}} - \bar{C}_n\|_{2p} \left(\|v_n^\ell - \bar{v}_n^\ell\|_{2p} + \|y_n^\ell - \bar{v}_n^\ell\|_{2p} \right). \end{aligned}$$

Plugging the moment bounds (3.25) into the right-hand side of the inequality (3.21) yields that $\|C_n^{\text{ML}} - \bar{C}_n^{\text{ML}}\|_{2p} \lesssim |\log(\varepsilon)|^{n-1} \varepsilon$, which in combination with Theorem 2 leads to $\|C_n^{\text{ML}} - C_n\|_{2p} \lesssim |\log(\varepsilon)|^{n-1} \varepsilon$. Therefore, summing the above and using (3.25) again for $p, 2p$

$$\begin{aligned} \sum_{\ell=0}^L \|\hat{v}_n^\ell - \bar{v}_n^\ell\|_p &\lesssim \sum_{\ell=0}^L \left\{ \|v_n^\ell - \bar{v}_n^\ell\|_p + \varepsilon \left(\|v_n^\ell - \bar{v}_n^\ell\|_{2p} + \|y_n^\ell - \bar{v}_n^\ell\|_{2p} \right) \right\} \\ &\lesssim |\log(\varepsilon)|^{n-1} \varepsilon \left(1 + \sum_{\ell=0}^L \|y_n^\ell - \bar{v}_n^\ell\|_{2p} \right) \\ &\lesssim |\log(\varepsilon)|^n \varepsilon. \end{aligned}$$

□

Induction is complete on the distance between the multilevel ensemble and its i.i.d. shadow in $L^p(\Omega; \mathcal{H})$, and we are finally ready to prove the main result.

Proof of Theorem 1. By Minkowski's inequality,

$$\begin{aligned} \|\hat{\mu}_n^{\text{ML}}(\varphi) - \hat{\mu}_n(\varphi)\|_p &\leq \|\hat{\mu}_n^{\text{ML}}(\varphi) - \hat{\mu}_n^{\text{ML}}(\varphi)\|_p + \|\hat{\mu}_n^{\text{ML}}(\varphi) - \hat{\mu}_n^L(\varphi)\|_p \\ &\quad + \|\hat{\mu}_n^L(\varphi) - \hat{\mu}_n(\varphi)\|_p, \end{aligned} \quad (3.27)$$

where $\hat{\mu}_n^{\text{ML}}$ denotes the empirical measure associated to the mean-field multilevel ensemble, and $\hat{\mu}_n^L$ denotes the probability measure associated to \hat{v}^L .

Before treating each term separately, we notice that the two first summands of the right-hand side of the inequality relates to the statistical error, whereas the last relates to the bias.

By the global Lipschitz continuity of the observable φ , Minkowski's inequality, and Lemma 6 the first term satisfies the following bound

$$\begin{aligned}
\|\hat{\mu}_n^{\text{ML}}(\varphi) - \hat{\mu}_n^{\text{ML}}(\varphi)\|_p &= \left\| \sum_{\ell=0}^L E_{M_\ell} [\varphi(\hat{v}_n^\ell) - \varphi(\hat{v}_n^{\ell-1}) - (\varphi(\hat{v}_n^\ell) - \varphi(\hat{v}_n^{\ell-1}))] \right\|_p \\
&\leq \sum_{\ell=0}^L \left(\|\varphi(\hat{v}_n^\ell) - \varphi(\hat{v}_n^{\ell-1})\|_p + \|\varphi(\hat{v}_n^{\ell-1}) - \varphi(\hat{v}_n^{\ell-1})\|_p \right) \\
&\leq C_\varphi \sum_{\ell=0}^L \left(\|\hat{v}_n^\ell - \hat{v}_n^{\ell-1}\|_p + \|\hat{v}_n^{\ell-1} - \hat{v}_n^{\ell-1}\|_p \right) \\
&\lesssim |\log(\varepsilon)|^n \varepsilon.
\end{aligned} \tag{3.28}$$

For the second summand of (3.27), notice that we can write $\hat{\mu}_n^L = \sum_{\ell=0}^L \hat{\mu}_n^\ell - \hat{\mu}_n^{\ell-1}$, where $\hat{\mu}_n^\ell$ is the measure associated to the level ℓ limiting process \hat{v}^ℓ and $\hat{\mu}_n^{-1} := 0$. Then, by virtue of Lemmas 3 and 7 and the global Lipschitz continuity of φ ,

$$\begin{aligned}
\|\hat{\mu}_n^{\text{ML}}(\varphi) - \hat{\mu}_n^L(\varphi)\|_p &\leq \sum_{\ell=0}^L \left\| E_{M_\ell} [\varphi(\hat{v}_n^\ell) - \varphi(\hat{v}_n^{\ell-1}) - \mathbb{E} [\varphi(\hat{v}_n^\ell) - \varphi(\hat{v}_n^{\ell-1})]] \right\|_p \\
&\leq c \sum_{\ell=0}^L M_\ell^{-1/2} \|\varphi(\hat{v}_n^\ell) - \varphi(\hat{v}_n^{\ell-1})\|_p \\
&\leq \tilde{c} \sum_{\ell=0}^L M_\ell^{-1/2} \|\hat{v}_n^\ell - \hat{v}_n^{\ell-1}\|_p \\
&\lesssim \sum_{\ell=0}^L M_\ell^{-1/2} h_\ell^{\beta/2} \lesssim \varepsilon.
\end{aligned} \tag{3.29}$$

Finally, the bias term in (3.27) satisfies

$$\|\hat{\mu}_n^L(\varphi) - \hat{\mu}_n(\varphi)\|_p = |\hat{\mu}_n^L(\varphi) - \hat{\mu}_n(\varphi)| = |\mathbb{E} [\varphi(\hat{v}_n^L) - \varphi(\hat{v}_n)]| \lesssim \varepsilon, \tag{3.30}$$

where the last step follows from the Lipschitz property and Lemma 3.

Inequalities (3.28), (3.29), and (3.30) together with inequality (3.27) complete the proof. \square

Theorem 1 shows the cost-to- ε performance of MLEnKF. The geometrically growing logarithmic penalty in the error (3.3) is disconcerting. The same penalty appears in the work [20], yet the numerical results there indicate a time-uniform rate of convergence, and this may be an artifact of the

rough bounds. We believe ergodicity of the MFEnKF process would allow us to obtain linear growth or even a uniform bound. There has been much recent work in this direction. The interested reader is referred to the works [9, 10, 35, 36, 42].

We conclude this section with a comparable result on the cost-to- ε performance of EnKF, showing that MLEnKF generally outperforms EnKF.

Theorem 3 (EnKF accuracy vs. cost). *Consider a globally Lipschitz continuous observable function $\varphi : \mathcal{H} \rightarrow \mathbb{R}$, and suppose Assumptions 1 and 2 hold. For a given $\varepsilon > 0$, let L and M be defined under the respective constraints $L = \lceil 2 \log_\kappa(\varepsilon^{-1})/\beta \rceil$ and $M \approx \varepsilon^{-2}$. Then, for any $n \in \mathbb{N}$ and $p \geq 2$,*

$$\|\hat{\mu}_n^{\text{MC}}(\varphi) - \hat{\mu}_n(\varphi)\|_p \lesssim \varepsilon, \quad (3.31)$$

where $\hat{\mu}_n^{\text{MC}}$ denotes the EnKF empirical measure, cf. equation (2.14), with particle evolution given by the EnKF predict and update formulae at resolution level L (i.e., with the numerical integrator Ψ^L in the prediction and projection operator \mathcal{P}_L in the update).

The computational cost of the EnKF estimator over the time sequence is bounded by

$$\text{Cost}(\text{EnKF}) \lesssim \varepsilon^{-2(1+d\gamma/\beta)}. \quad (3.32)$$

Sketch of proof. By Minkowski's inequality,

$$\begin{aligned} \|\hat{\mu}_n(\varphi) - \hat{\mu}_n^{\text{MC}}(\varphi)\|_p &\leq \|\hat{\mu}_n(\varphi) - \hat{\mu}_n^L(\varphi)\|_p + \|\hat{\mu}_n^L(\varphi) - \hat{\mu}_n^{\text{MC}}(\varphi)\|_p \\ &\quad + \|\hat{\mu}_n^{\text{MC}}(\varphi) - \hat{\mu}_n^{\text{MC}}(\varphi)\|_p =: I + II + III, \end{aligned}$$

where $\hat{\mu}_n^{\text{MC}}$ denotes the empirical measure associated to the EnKF ensemble $\{\hat{v}_{n,i}^L\}_{i=1}^M$ and $\hat{\mu}_n^L$ denotes the empirical measure associated to \hat{v}_n^L . It follows by inequality (3.30) that $I \lesssim \varepsilon$.

For the second term, note that (3.1) guarantees the existence of a positive scalar C_φ such that $|\varphi(x)| \leq C_\varphi(1 + \|x\|_{\mathcal{H}})$. Since $\hat{v}_n^L \in L^p(\Omega; \mathcal{H})$ for any $n \in \mathbb{N}$ and $p \geq 2$, it follows by Lemma 7 (on the Hilbert space \mathcal{R}_1) that

$$II \leq \|E_M[\varphi(\hat{v}_n^L)] - \mathbb{E}[\varphi(\hat{v}_n^L)]\|_p \leq M^{-1/2} C_\varphi \|\hat{v}_n^L\|_p \lesssim \varepsilon.$$

For the last term, let us first assume that for any $p \geq 2$ and $n \in \mathbb{N}$,

$$\|\hat{v}_{n,1}^L - \hat{v}_{n,1}^L\|_p \lesssim \varepsilon, \quad (3.33)$$

for the single particle dynamics $\hat{v}_{n,1}^L$ and $\hat{v}_{n,1}^L$ respectively associated to the EnKF ensemble $\{\hat{v}_{n,i}^L\}_{i=1}^M$ and the mean-field EnKF ensemble $\{\hat{v}_{n,i}^L\}_{i=1}^M$. Then the global Lipschitz continuity of φ , the fact that $\hat{v}_{n,1}^L, \hat{v}_{n,1}^L \in L^p(\Omega; \mathcal{H})$ for any $n \in \mathbb{N}$ and $p \geq 2$, and Minkowski's inequality yield that

$$III = \|E_M[\varphi(\hat{v}_n^L) - \varphi(\hat{v}_n^L)]\|_p \leq C_\varphi \|\hat{v}_n^L - \hat{v}_n^L\|_p \lesssim \varepsilon.$$

All that remains is to verify (3.33), but we omit this verification as it can be done by similar steps as in the proof of inequality (3.23). \square

4. A CONCRETE EXAMPLE

Consider the stochastic heat equation, given abstractly as follows:

$$du = -Audt + BdW, \quad u(0) \sim N(0, C_0), \quad (4.1)$$

where A is the abstract representation of $(-\Delta)$ acting on the space $\mathcal{H} := \{u \in L^2(D); \int_D u(x)dx = 0\}$, $B = A^{-b}$ for some $b \geq 0$ and $C_0 = A^{-a}$ for some $a \geq 0$. Let $D = [-\pi, \pi]^d$. The standard Sobolev spaces are defined as follows.

Definition 2. \mathcal{H}^s is defined as the space $\{u \in \mathcal{H}; \langle u, A^s u \rangle_{\mathcal{H}} < \infty\}$, with the associated norm $\|u\|_{\mathcal{H}^s} = \langle u, A^s u \rangle_{\mathcal{H}}$.

Consider the Fourier basis $\{\phi_k\}_{k=-\infty}^{\infty}$ such that $\phi_k(x) = e^{-ik \cdot x}$, $i = \sqrt{-1}$, and for $u \in \mathcal{H}$ one has the expansion $u = \sum_{k=-\infty}^{\infty} u_k \phi_k(x)$ subject to reality constraint $u_{-k} = u_k^*$, and with $u_0 = 0$. One has spectral expansions $A = \sum_{k=-\infty}^{\infty} |k|^2 \phi_k \otimes \phi_k$, $B = \sum_{k=-\infty}^{\infty} b_k \phi_k \otimes \phi_k$, and $C = \sum_{k=-\infty}^{\infty} c_k \phi_k \otimes \phi_k$. The solution for u_k , $k \geq 1$, is given analytically as

$$u_k(t) = e^{-k^2 t} u_k(0) + \xi_k(t), \quad \xi_k(t) \sim N \left[0, \frac{b_k^2}{2k^2} (1 - e^{-2k^2 t}) \right] \perp u_k(0). \quad (4.2)$$

For observation increment τ , the observations are taken as

$$y_n = Hu(\tau n) + \eta_n, \quad \eta_n \sim N(0, \Gamma) \text{ i.i.d. } \perp u(0), \xi_k(\tau n) \forall k. \quad (4.3)$$

The observation operator may be taken as $Hu = [H_1(u), \dots, H_m(u)]^\top$, where $H_i(u) = \int u(x) \psi_i(x) dx$ for some $\psi_i \in \mathcal{H}$. Notice the model is non-trivial as correlations will arise from the update unless $\psi_i = \phi_k$ for some k .

Note that for this simple Gaussian model one simply requires that $u \in L^2(\Omega; \mathcal{H})$, since all other moments are controlled by the variance. Indeed if $\mathbb{E}\|u\|_{\mathcal{H}}^2 < \infty$, then $u \in L^p(\Omega; \mathcal{H})$ for all p and $u \in \mathcal{H}$ almost surely.

Following from (4.2) define

$$\Psi(u) := \sum_{k=0}^{\infty} \left(e^{-k^2} u_k + \xi_k \right) \phi_k, \quad \xi_k \sim N \left[0, \frac{b_k^2}{2k^2} (1 - e^{-2k^2}) \right], \quad (4.4)$$

where $u_k = \langle \phi_k, u \rangle$. Notice that the regularity of $\Psi(u)$ does not depend on u at all (assuming it is not exponentially rough). Indeed by the assumed form of B , one has $b_k = \mathcal{O}(k^{-2b})$. Notice

$$\mathbb{E}\|\Psi(u)\|_{\mathcal{H}^s}^2 = \sum_{k=1}^{\infty} k^{2s} \left(e^{-2k^2} \mathbb{E}u_k^2 + \frac{1}{2k^{2(2b+1)}} (1 - e^{-2k^2}) \right).$$

Therefore, $\Psi(u) \in \mathcal{H}^s$ for any $s < 2b + 1 - d/2$.

Indeed, $\Psi(u)$ is Gaussian with a *smoothing* covariance C , such that for $u \in \mathcal{H}^s$, one has $Cu \in \mathcal{H}^{s+2b+1}$. Assuming that $H : \mathcal{H} \rightarrow \mathbb{R}^m$ is defined by \mathcal{H} inner products, then $H^* : \mathbb{R}^m \rightarrow \mathcal{H}^* = \mathcal{H}$. Hence the Kalman gain $K : \mathbb{R}^m \rightarrow \mathcal{H}^{2b+1} \subset \mathcal{H}$, following from the form of (2.25).

For a concrete example, let $d = 1$ and $b = 0$. Then $\mathcal{H} := L^2(D)$, and $u_n \in L^p(\Omega; \mathcal{H})$ for all $n \in \mathbb{N}$ and $p \geq 2$. Assume \mathcal{P}_ℓ is the projection onto 2^ℓ

Fourier modes. The k^{th} mode is given by $u_{n,k} = N(u_{n-1,k}e^{-k^2}, \sigma_k^2)$, where $\sigma_k^2 = \mathcal{O}(k^{-2})$. This in turn induces a rate of convergence of

$$\|(I - \mathcal{P}_\ell)u_n\|_{L^2(\Omega; \mathcal{H})} = \mathcal{O}\left(\left(\sum_{\{k > 2^\ell\}} \sigma_k^2\right)^{1/2}\right) = \mathcal{O}\left(2^{-\ell/2}\right),$$

as $\ell \rightarrow \infty$. Higher moments follow from Gaussianity, with a p -dependent constant. The other assumptions are easily verified as well.

5. CONCLUSION

An extension of the recent work [20] to spatially extended models is presented here, using a hierarchical decomposition based on the spatial resolution parameter. The proof follows closely that of [20], except with the important extension to infinite-dimensions. It is shown that an optimality rate similar to vanilla MLMC can extend to the case of sequential inference using EnKF for spatial models as well. One may therefore expect that value can be leveraged, for a fixed computational cost, by spreading work across a multilevel ensemble associated to models of multiple spatial resolutions rather than restricting to an ensemble associated only to the finest resolution model and using one very small ensemble. This has potential for broad impact across application areas in which there has been a recent explosion of interest in EnKF, for example weather prediction and subsurface exploration.

APPENDIX A. MARCINKIEWICZ–ZYGmund INEQUALITIES FOR HILBERT SPACES

For closing the proof of Lemma 4 we make use a couple of lemmas extending the Marcinkiewicz–Zygmund inequality to separable Banach spaces.

Lemma 7. [29, Theorem 5.2] *Let $2 \leq p < \infty$ and $X_i \in L^p(\Omega; \mathcal{H})$ be i.i.d. samples of $X \in L^p(\Omega; \mathcal{H})$. Then*

$$\|E_M[X] - \mathbb{E}[X]\|_{L^p(\Omega; \mathcal{H})} \leq \frac{c_p}{\sqrt{M}} \|X - \mathbb{E}[X]\|_{L^p(\Omega; \mathcal{H})} \quad (\text{A.1})$$

where c_p only depends on p .

Proof. Let r_1, r_2, \dots denote a sequence of real-valued i.i.d. random variables with $P(r_i = \pm 1) = 1/2$. A Banach space \mathcal{K} is said to be of R-type q if there exists a $c > 0$ such that for every $\bar{n} \in \mathbb{N}$ and for all (deterministic) $x_1, x_2, \dots, x_{\bar{n}} \in \mathcal{K}$,

$$\mathbb{E}\left[\left\|\sum_{i=1}^{\bar{n}} r_i x_i\right\|_{\mathcal{K}}\right] \leq c \left(\sum_{i=1}^{\bar{n}} \|x_i\|_{\mathcal{K}}^q\right)^{1/q}.$$

It is clear that all Hilbert spaces (and for our interest \mathcal{H} , in particular) are of R-type 2, since their norms are induced by an inner product. Following the proofs of [43, Proposition 2.1 and Corollary 2.1], we introduce the symmetrization $\tilde{X}_i := (X_i - X'_i)$ and derive that

$$\begin{aligned} \mathbb{E} \left[\left\| \sum_{i=1}^{\bar{n}} X_i - \mathbb{E}[X] \right\|_{\mathcal{H}}^p \right] &\leq \mathbb{E} \left[\left\| \sum_{i=1}^{\bar{n}} \tilde{X}_i \right\|_{\mathcal{H}}^p \right] = \mathbb{E} \left[\left\| \sum_{i=1}^{\bar{n}} r_i \tilde{X}_i \right\|_{\mathcal{H}}^p \right] \\ &\leq c \mathbb{E} \left[\left(\sum_{i=1}^{\bar{n}} \|\tilde{X}_i\|_{\mathcal{H}}^2 \right)^{p/2} \right] \leq c 2^p \mathbb{E} \left[\left(\sum_{i=1}^{\bar{n}} \|X_i - \mathbb{E}[X]\|_{\mathcal{H}}^2 \right)^{p/2} \right]. \end{aligned}$$

And by another application of Hölder's inequality,

$$\begin{aligned} \mathbb{E} \left[\left\| \sum_{i=1}^M \frac{X_i - \mathbb{E}[X]}{M} \right\|_{\mathcal{H}}^p \right] &\leq \hat{c} M^{-p} \mathbb{E} \left[\left(\sum_{i=1}^M \|X_i - \mathbb{E}[X]\|_{\mathcal{H}}^2 \right)^{p/2} \right] \\ &\leq \hat{c} M^{-p/2} \mathbb{E} [\|X - \mathbb{E}[X]\|_{\mathcal{H}}^p]. \end{aligned}$$

□

Lemma 8. *Suppose $X, Y \in L^p(\Omega; \mathcal{H})$, $p \geq 2$. Then, for $1 \leq r, s \leq \infty$ satisfying $1/r + 1/s = 1$, it holds that*

$$\|\text{Cov}_M[X, Y] - \text{Cov}[X, Y]\|_{L^p(\Omega; \mathcal{H} \otimes \mathcal{H})} \leq \frac{c}{\sqrt{M}} \|X\|_{L^{pr}(\Omega; \mathcal{H})} \|Y\|_{L^{ps}(\Omega; \mathcal{H})} \quad (\text{A.2})$$

where the upper bound for the constant $c = \frac{M}{M-1} \left(2c_p + \frac{c_{pr}c_{ps} + 1}{\sqrt{M}} \right)$ only depends on r, s and p .

Proof. Since $\text{Cov}[X, Y] = \text{Cov}[X - \mathbb{E}[X], Y - \mathbb{E}[Y]]$ and $\text{Cov}_M[X, Y] = \text{Cov}_M[X - \mathbb{E}[X], Y - \mathbb{E}[Y]]$, cf. (2.12), we may without loss of generality assume that $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. Using Minkowski's inequality

$$\begin{aligned} &\frac{M-1}{M} \|\text{Cov}_M[X, Y] - \text{Cov}[X, Y]\|_p \\ &\leq \|E_M[X \otimes Y] - \mathbb{E}[X \otimes Y]\|_p + \|E_M[X] \otimes E_M[Y]\|_p + \frac{1}{M} \|\mathbb{E}[X \otimes Y]\|_{\mathcal{H} \otimes \mathcal{H}}. \end{aligned} \quad (\text{A.3})$$

We estimate the three terms in the right-hand side separately. Estimate (A.1) and Hölder's inequality yield

$$\begin{aligned} \|E_M[X \otimes Y] - \mathbb{E}[X \otimes Y]\|_p &\leq \frac{c_p}{\sqrt{M}} \|X \otimes Y - \mathbb{E}[X \otimes Y]\|_p \\ &\leq \frac{2c_p}{\sqrt{M}} \|X \otimes Y\|_p \leq \frac{2c_p}{\sqrt{M}} \|X\|_{pr} \|Y\|_{ps}. \end{aligned}$$

Similarly, since $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ by assumption, we obtain by (A.1) and Hölder's inequality

$$\|E_M[X] \otimes E_M[Y]\|_p \leq \|E_M[X]\|_{pr} \|E_M[Y]\|_{ps} \leq \frac{c_{pr} c_{ps}}{M} \|X\|_{pr} \|Y\|_{ps}. \quad (\text{A.4})$$

And, finally, for the last term

$$\frac{1}{M} \|\mathbb{E}[X \otimes Y]\|_{\mathcal{H} \otimes \mathcal{H}} \leq \frac{1}{M} \|X \otimes Y\|_{L^1(\Omega, \mathcal{H} \otimes \mathcal{H})} \leq \frac{1}{M} \|X\|_{L^{pr}(\Omega; \mathcal{H})} \|Y\|_{L^{ps}(\Omega; \mathcal{H})}.$$

□

Acknowledgements Research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST). HH was additionally supported by Norges Forskningsråd, research project 214495 LIQCRY. KJHL was additionally supported by an ORNL LDRD Strategic Hire grant.

REFERENCES

- [1] A. Bain and D. Crisan. *Fundamentals of Stochastic Filtering*. Springer, 2009.
- [2] Alexandros Beskos, Ajay Jasra, Kody Law, Raul Tempone, and Yan Zhou. Multi-level sequential Monte Carlo samplers. *To appear in Stochastic Processes and their Applications* <http://dx.doi.org/10.1016/j.spa.2016.08.004>.
- [3] Dirk Blömker, Kody Law, Andrew M Stuart, and Konstantinos C Zygalakis. Accuracy and stability of the continuous-time 3DVAR filter for the Navier–Stokes equation. *Nonlinearity*, 26(8):2193, 2013.
- [4] CEA Brett, Kei Fong Lam, KJH Law, DS McCormick, MR Scott, and AM Stuart. Accuracy and stability of filters for dissipative PDEs. *Physica D: Nonlinear Phenomena*, 245(1):34–45, 2013.
- [5] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly weather review*, 126(6):1719–1724, 1998.
- [6] Kai Lai Chung. *A course in probability theory*. Academic Press, Inc., San Diego, CA, third edition, 2001.
- [7] Pierre Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.
- [8] Pierre Del Moral, Ajay Jasra, Kody Law, and Yan Zhou. Multilevel sequential Monte Carlo samplers for normalizing constants. *arXiv preprint arXiv:1603.01136*, 2016.
- [9] Pierre Del Moral, Aline Kurtzmann, and Julian Tugaut. On the stability and the uniform propagation of chaos of extended ensemble Kalman-Bucy filters. *arXiv preprint arXiv:1606.08256*, 2016.
- [10] Pierre Del Moral and Julian Tugaut. On the stability and the uniform propagation of chaos properties of ensemble Kalman-Bucy filters. *arXiv preprint arXiv:1605.09329*, 2016.
- [11] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and computing*, 10(3):197–208, 2000.
- [12] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans (1978–2012)*, 99(C5):10143–10162, 1994.
- [13] Geir Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4):343–367, 2003.

- [14] Aseel Farhat, Evelyn Lunasin, and Edriss S Titi. Data assimilation algorithm for 3D Bénard convection in porous media employing only temperature measurements. *arXiv preprint arXiv:1506.08678*, 2015.
- [15] M. B. Giles and L. Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Ann. Appl. Probab.*, 24(4):1585–1620, 2014.
- [16] Alastair Gregory, CJ Cotter, and Sebastian Reich. Multilevel ensemble transform particle filtering. *SIAM Journal on Scientific Computing*, 38(3):A1317–A1338, 2016.
- [17] Kevin Hayden, Eric Olson, and Edriss S Titi. Discrete data assimilation in the Lorenz and 2d Navier–Stokes equations. *Physica D: Nonlinear Phenomena*, 240(18):1416–1425, 2011.
- [18] Stefan Heinrich. Multilevel Monte Carlo methods. In *Large-scale scientific computing*, pages 58–67. Springer, 2001.
- [19] Viet Ha Hoang, Christoph Schwab, and Andrew M Stuart. Complexity analysis of accelerated mcmc methods for Bayesian inversion. *Inverse Problems*, 29(8):085010, 2013.
- [20] Håkon Hoel, Kody Law, and Raul Tempone. Multilevel ensemble Kalman filter. *SIAM Journal of Numerical Analysis*, 54(3):1813–1839, 2016.
- [21] Ajay Jasra, Kengo Kamatani, Kody JH Law, and Yan Zhou. Multilevel particle filter. *arXiv preprint arXiv:1510.04977*, 2015.
- [22] Ajay Jasra, Kody Law, and Yan Zhou. Forward and inverse uncertainty quantification using multilevel Monte Carlo algorithms for an elliptic nonlocal equation. *arXiv preprint arXiv:1603.06381*, 2016.
- [23] A.H. Jazwinski. *Stochastic processes and filtering theory*, volume 63. Academic Pr, 1970.
- [24] Rudolph Emil Kalman et al. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.
- [25] E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge, 2003.
- [26] David Kelly, Andrew J Majda, and Xin T Tong. Concrete ensemble Kalman filters with rigorous catastrophic filter divergence. *Proceedings of the National Academy of Sciences*, 112(34):10589–10594, 2015.
- [27] DTB Kelly, KJH Law, and Andrew M Stuart. Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time. *Nonlinearity*, 27(10):2579, 2014.
- [28] C Ketelsen, R Scheichl, and AL Teckentrup. A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow. *arXiv preprint arXiv:1303.7343*, 2013.
- [29] Evan Kwiatkowski and Jan Mandel. Convergence of the square root ensemble Kalman filter in the large ensemble limit. *arXiv preprint arXiv:1404.4093*, 2014.
- [30] KJH Law, AM Stuart, and KC Zygalakis. Data assimilation: A mathematical introduction. *Springer Texts in Applied Mathematics*, 2015.
- [31] Kody JH Law, Abhishek Shukla, and Andrew M Stuart. Analysis of the 3DVAR filter for the partially observed Lorenz’63 model. *Discrete and Continuous Dynamical Systems*, 34(3):1061–1078, 2013.
- [32] Kody JH Law, Hamidou Tembine, and Raul Tempone. Deterministic mean-field ensemble Kalman filtering. *SIAM Journal on Scientific Computing*, 38(3):A1251–A1279, 2016.
- [33] François Le Gland, Valérie Monbet, Vu-Duc Tran, et al. Large sample asymptotics for the ensemble Kalman filter. *The Oxford Handbook of Nonlinear Filtering*, pages 598–631, 2011.
- [34] David G Luenberger. *Optimization by vector space methods*. John Wiley & Sons, 1968.

- [35] Andrew J Majda and Xin T Tong. Rigorous accuracy and robustness analysis for two-scale reduced random Kalman filters in high dimensions. *arXiv preprint arXiv:1606.09087*, 2016.
- [36] Andrew J Majda and Xin T Tong. Robustness and accuracy of finite ensemble Kalman filters in large dimensions. *arXiv preprint arXiv:1606.09321*, 2016.
- [37] Jan Mandel, Loren Cobb, and Jonathan D Beezley. On the convergence of the ensemble Kalman filter. *Applications of Mathematics*, 56(6):533–541, 2011.
- [38] Eric Olson and Edriss S Titi. Determining modes for continuous data assimilation in 2d turbulence. *Journal of statistical physics*, 113(5-6):799–840, 2003.
- [39] Oliver Pajonk, Bojana V Rosić, Alexander Litvinenko, and Hermann G Matthies. A deterministic filter for non-Gaussian Bayesian estimation applications to dynamical system estimation with noisy measurements. *Physica D: Nonlinear Phenomena*, 241(7):775–788, 2012.
- [40] Tzyh-Jong Tarn and YONA Rasis. Observers for nonlinear stochastic systems. *IEEE Transactions on Automatic Control*, 21(4):441–448, 1976.
- [41] Xin T Tong, Andrew J Majda, and David Kelly. Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation. *arXiv preprint arXiv:1507.08319*, 2015.
- [42] Xin T Tong, Andrew J Majda, and David Kelly. Nonlinear stability and ergodicity of ensemble based Kalman filters. *Nonlinearity*, 29(2):657, 2016.
- [43] Wojbor A. Woyczyński. On Marcinkiewicz-Zygmund laws of large numbers in Banach spaces and related rates of convergence. *Probab. Math. Statist.*, 1(2):117–131 (1981), 1980.